Variational Methods for Latent Variable Problems

Ryan Giordano (for Johns Hopkins Biostats BLAST working group)

Oct, 2021

Massachusetts Institute of Technology

Outline

Outline for today:

- Some examples of latent variable models
- A template: The Neyman-Scott "paradox" and marginalization
- Bayesian versus frequentist approaches to marginalization
- The classical EM algorithm (in brief)

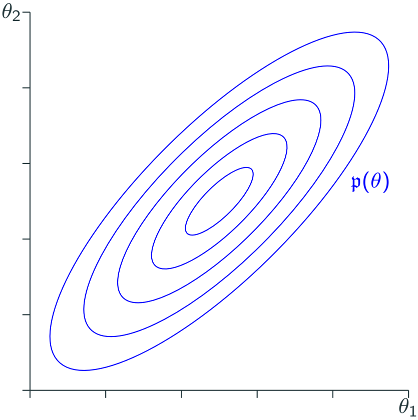
Next week, we will build on these ideas to present more general variational inference.

KL divergence exercises

$$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$$

$$\mathcal{Q} = \{\text{All bivariate normals}\}$$

What is $\mathfrak{q}^*(\theta) = \underset{\mathfrak{q} \in \mathcal{Q}}{\operatorname{argmin}} \, \mathrm{KL}\left(\mathfrak{q}(\theta) || \mathfrak{p}(\theta)\right)$?

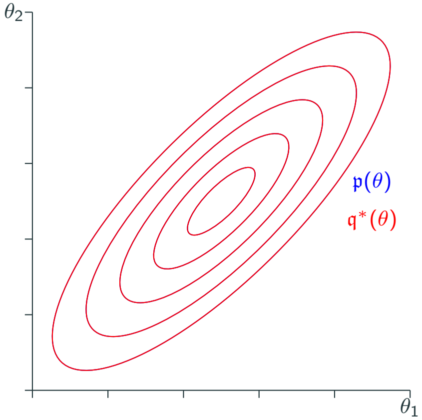


KL divergence exercises

$$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$$

$$\mathcal{Q} = \{\text{All bivariate normals}\}$$

What is $\mathfrak{q}^*(\theta) = \underset{\mathfrak{q} \in \mathcal{Q}}{\operatorname{argmin}} \operatorname{KL}\left(\mathfrak{q}(\theta) || \mathfrak{p}(\theta)\right)$?

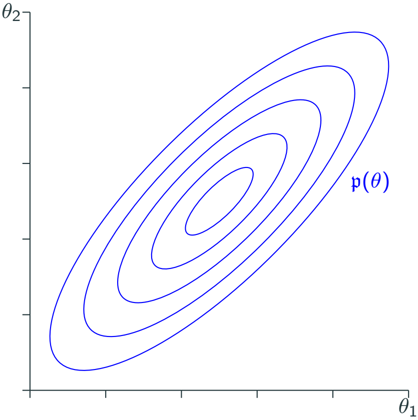


KL divergence exercises

$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$

$\mathcal{Q} = \{\text{Independent bivariate normals}\}$

What is $\mathfrak{q}^*(\theta) = \underset{\mathfrak{q} \in \mathcal{Q}}{\operatorname{argmin}} \operatorname{KL}(\mathfrak{q}(\theta) || \mathfrak{p}(\theta))$?

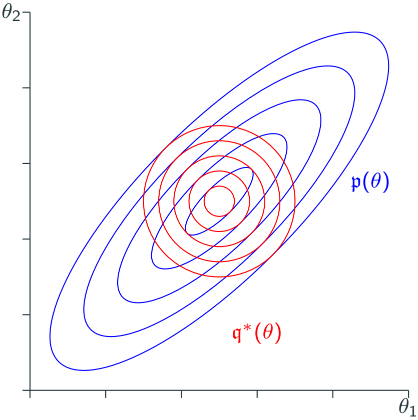


KL divergence exercises

$$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$$

$$\mathcal{Q} = \{\text{Independent bivariate normals}\}$$

What is $\mathfrak{q}^*(\theta) = \underset{\mathfrak{q} \in \mathcal{Q}}{\operatorname{argmin}} \, \mathrm{KL}\left(\mathfrak{q}(\theta) || \mathfrak{p}(\theta)\right)$?

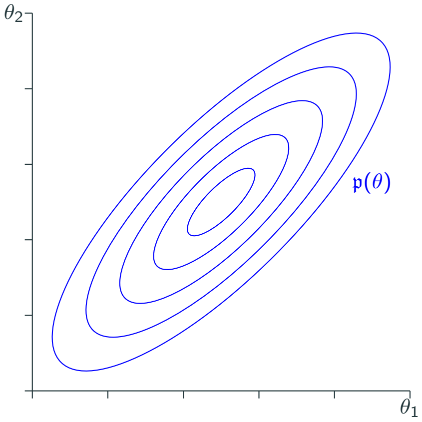


KL divergence exercises

$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$

$\mathcal{Q} = \{\text{Independent bivariate normals}\}$

What is $\mathfrak{q}^*(\theta) = \underset{\mathfrak{q} \in \mathcal{Q}}{\operatorname{argmin}} \, \mathrm{KL}\left(\mathfrak{p}(\theta) || \mathfrak{q}(\theta)\right)$?

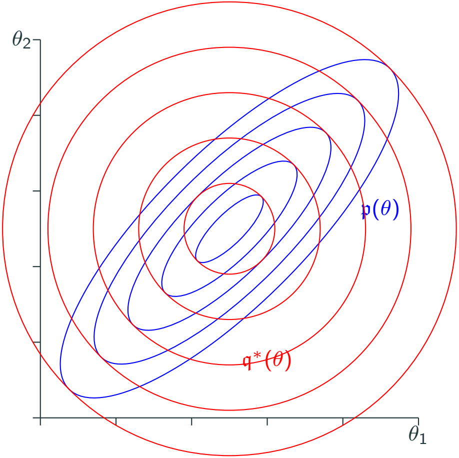


KL divergence exercises

$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$

$\mathcal{Q} = \{\text{Independent bivariate normals}\}$

What is $\mathfrak{q}^*(\theta) = \operatorname*{argmin}_{\mathfrak{q} \in \mathcal{Q}} \mathrm{KL}\left(\mathfrak{p}(\theta) || \mathfrak{q}(\theta)\right)$?



KL divergence exercises

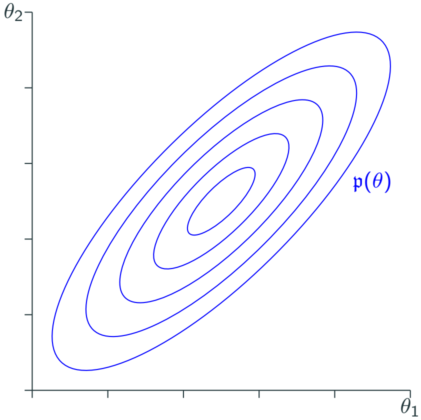
Recall that

$$\mathrm{KL}\left(\mathfrak{q}(\theta)||\mathfrak{p}(\theta)\right) =$$
$$- \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{p}(\theta)\right] + \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{q}(\theta)\right]$$

$$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$$

$$\mathcal{Q} = \{\text{Bivariate normals}\}$$

What is $\mathfrak{q}^*(\theta) =$
$$\underset{\mathfrak{q}\in\mathcal{Q}}{\mathrm{argmin}}\left(- \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{p}(\theta)\right] + \cancel{\underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{q}(\theta)\right]}\right)?$$



KL divergence exercises

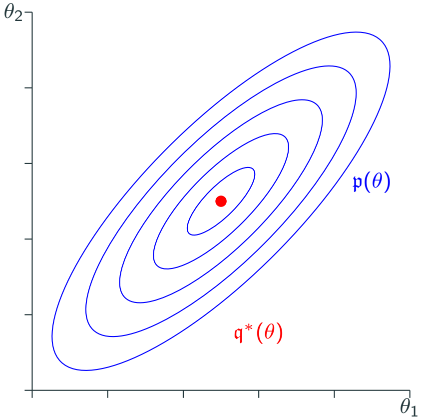
Recall that

$$\mathrm{KL}\left(\mathfrak{q}(\theta)||\mathfrak{p}(\theta)\right) = -\underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{p}(\theta)\right] + \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{q}(\theta)\right]$$

$$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$$

$$\mathcal{Q} = \{\text{Bivariate normals}\}$$

What is $\mathfrak{q}^*(\theta) = \underset{\mathfrak{q}\in\mathcal{Q}}{\mathrm{argmin}}\left(-\underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{p}(\theta)\right] + \cancel{\underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{q}(\theta)\right]}\right)$?



KL divergence exercises

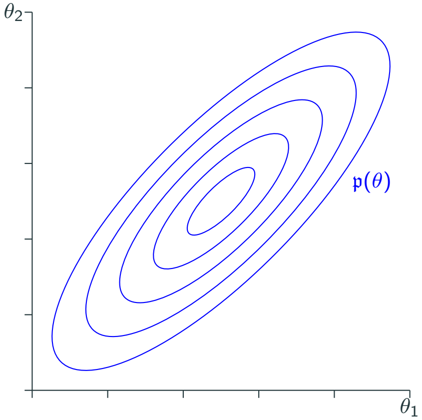
Recall that

$$\mathrm{KL}\left(\mathfrak{q}(\theta)||\mathfrak{p}(\theta)\right) = - \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{p}(\theta)\right] + \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{q}(\theta)\right]$$

$$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$$

$$\mathcal{Q} = \{\text{Bivariate normals}\}$$

What is $\mathfrak{q}^*(\theta) = \underset{\mathfrak{q}\in\mathcal{Q}}{\operatorname{argmin}}\left(- \cancel{\underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{p}(\theta)\right]} + \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{q}(\theta)\right]\right)$?



KL divergence exercises

Recall that

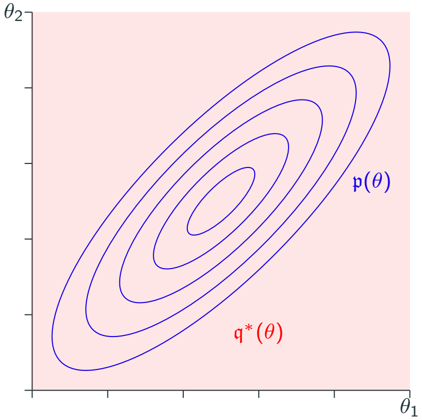
$$\mathrm{KL}\left(\mathfrak{q}(\theta)||\mathfrak{p}(\theta)\right) = - \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{p}(\theta)\right] + \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{q}(\theta)\right]$$

$$\mathfrak{p}(\theta) = \text{Correlated bivariate normal}$$

$$\mathcal{Q} = \{\text{Bivariate normals}\}$$

What is $\mathfrak{q}^*(\theta) =$

$$\underset{\mathfrak{q}\in\mathcal{Q}}{\mathrm{argmin}}\left(- \cancel{\underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{p}(\theta)\right]} + \underset{\mathfrak{q}(\theta)}{\mathbb{E}}\left[\log \mathfrak{q}(\theta)\right]\right)?$$



Conclusions